

Global mapping of c-Myc binding sites and target gene networks in human B cells

Karen I. Zeller*, XiaoDong Zhao[†], Charlie W. H. Lee[†], Kuo Ping Chiu[†], Fei Yao[†], Jason T. Yustein*, Hong Sain Ooi[†], Yuriy L. Orlov[†], Atif Shahab[‡], How Choong Yong[‡], YuTao Fu[§], Zhiping Weng[§], Vladimir A. Kuznetsov[†], Wing-Kin Sung[†], Yijun Ruan[†], Chi V. Dang*[¶], and Chia-Lin Wei*[¶]

*Department of Medicine and The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, The Johns Hopkins University School of Medicine, Baltimore, MD 21205; [†]Genome Institute of Singapore, Singapore 138672; [‡]Bioinformatics Institute, Singapore 138671; and [§]Bioinformatics Program, Boston University, Boston, MA 02115

Edited by Robert N. Eisenman, Fred Hutchinson Cancer Research Center, Seattle, WA, and approved September 27, 2006 (received for review May 18, 2006)

The protooncogene *MYC* encodes the c-Myc transcription factor that regulates cell growth, cell proliferation, cell cycle, and apoptosis. Although deregulation of *MYC* contributes to tumorigenesis, it is still unclear what direct Myc-induced transcriptomes promote cell transformation. Here we provide a snapshot of genome-wide, unbiased characterization of direct Myc binding targets in a model of human B lymphoid tumor using ChIP coupled with pair-end ditag sequencing analysis (ChIP-PET). Myc potentially occupies >4,000 genomic loci with the majority near proximal promoter regions associated frequently with CpG islands. Using gene expression profiles with ChIP-PET, we identified 668 direct Myc-regulated gene targets, including 48 transcription factors, indicating that Myc is a central transcriptional hub in growth and proliferation control. This first global genomic view of Myc binding sites yields insights of transcriptional circuitries and cis regulatory modules involving Myc and provides a substantial framework for our understanding of mechanisms of Myc-induced tumorigenesis.

human genome | chromatin immunoprecipitation | pair-end ditagging | oncogene | tumorigenesis

The protooncogene *MYC* encodes a transcription factor, c-Myc (herein termed Myc), that regulates cell size, cell proliferation, and apoptosis (1, 2). Normal expression of *MYC* is exquisitely regulated, such that mitogens induce its expression when normal cells are recruited into the cell cycle (1). Conversely, cellular quiescence and differentiation dramatically diminish *MYC* expression. By contrast, cancer cells bear genetic alterations that deregulate *MYC* expression, and constitutive expression of Myc is central to its transforming activity. Myc is a basic helix-loop-helix leucine zipper protein that dimerizes with Max to bind the DNA sequence 5'-CACGTG-3', known as an E box, and activates transcription (3). Myc also represses transcription through an interaction with Miz-1 or through other elements at core promoters (4); however, the mechanisms associated with the latter are not well understood. The transcriptional activity of Myc is crucial for its ability to cause malignant transformation, because transcriptionally defective *MYC* alleles have diminished transforming potential (5).

To elucidate the role of *MYC* in tumorigenesis and development, many efforts have focused on identifying Myc target genes and how transcriptional alteration of these targets leads to cell size increase, cell-cycle progression, apoptosis, or abrogation of cell differentiation (6). To date, there are >1,500 genes found to be Myc-responsive genes and compiled in the Myc target gene database (www.mycancergene.org) (7). Recently, high-throughput expression profiling such as microarray (8, 9) and serial analysis of gene expression (SAGE) (10) have been adopted to identify hundreds of Myc-responsive genes. Because most expression studies were limited by the paucity of target validation by quantitative PCR (qPCR) and were unable to definitively distinguish between direct and indirect targets, only

a minority of the Myc-responsive genes have been implicated as direct target genes.

ChIP is a powerful technique to identify direct target genes via isolating DNA fragments bound by proteins (11). When coupled with microarray detection method (ChIP-chip) or ChIP-qPCR (qPCR of ChIP products), direct Myc binding loci on complex genomes could be identified (12–15). However, the available ChIP studies only focused on a few highly selective features and regions of the human genome. Notwithstanding the limitations of current studies, they collectively suggest that Myc could regulate up to 10–15% of all genes. Hence, it is critically important to define a direct Myc target transcriptome using well defined, tractable experimental system and a method that permits global mapping of Myc binding sites. Recently, we have developed an unbiased whole-genome mapping strategy to identify transcription factor binding sites, called ChIP-PET (ChIP coupled with pair-end ditagging) (16). With ChIP-PET, ChIP DNA fragments are first cloned, and then the 36-bp paired 5' and 3' tag for each of the cloned fragments are extracted. These pair-end ditags (PETs) are further concatenated for efficient sequencing and accurate mapping to reference genome for demarcation of ChIP DNA fragments. The overlapping of PET-inferred ChIP DNA fragments has been proven an effective readout to reflect enriched DNA loci, which identify transcription factor binding sites. The utility of ChIP-PET approach had been demonstrated in mapping human genomic p53 binding sites (16) and elucidating the transcription networks of Oct4 and Nanog in the mouse genome (17).

To further delineate the Myc transcriptional network, we performed a Myc ChIP-PET study with a model human B cell line P493 coupled with gene expression data and computational analysis. Herein we present a whole-genome Myc binding profile and elucidate Myc directed transcriptome and transcription regulatory networks.

Results

Global Mapping of Myc Binding Sites in Human P493 B Cells. To identify Myc binding sites in a human genome, a human B cell

Author contributions: K.I.Z. and X.D.Z. contributed equally to this work; C.V.D. and C.-L.W. designed research; K.I.Z., X.D.Z., F.Y., and J.T.Y. performed research; C.W.H.L., A.S., H.C.Y., C.V.D., and C.-L.W. contributed new reagents/analytic tools; K.I.Z., X.D.Z., C.W.H.L., K.P.C., H.S.O., Y.L.O., Y.T.F., Z.W., V.A.K., W.-K.S., C.V.D., and C.-L.W. analyzed data; and K.I.Z., X.D.Z., V.A.K., Y.R., C.V.D., and C.-L.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Freely available online through the PNAS open access option.

Abbreviations: qPCR, quantitative PCR; PET, pair-end ditag; EBF, early B cell factor.

Data deposition: The data presented in this paper have been deposited with the ENCODE project at UCSC, <http://genome.ucsc.edu/ENCODE> (track GIS c-Myc P493).

[¶]To whom correspondence may be addressed. E-mail: cvdang@jhmi.edu or weic@gis-star.edu.sg.

© 2006 by The National Academy of Sciences of the USA

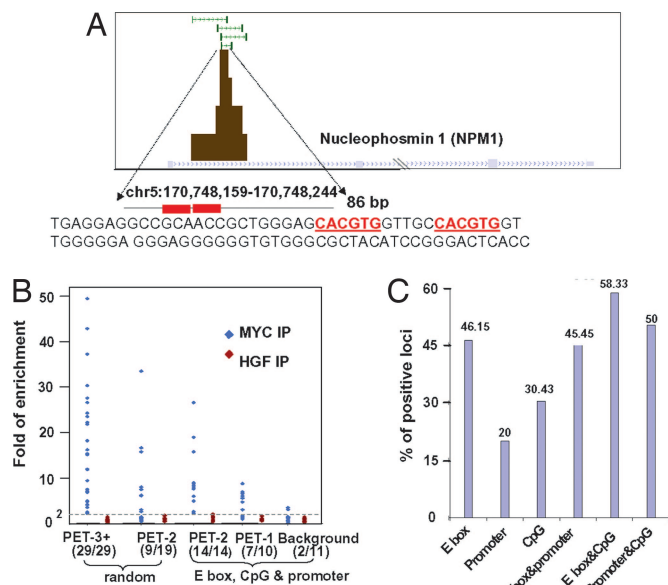


Fig. 2. Validation of PET clusters as reliable readouts of Myc binding. (A) A known Myc binding site in the first intron of *NPM1* is localized by a PET-4 cluster. This 86-bp overlap region contains two tandem canonical E boxes. (B) Myc ChIP DNA (blue) and control HGF ChIP DNA (red) from an independent ChIP experiment were subjected to ChIP-qPCR to validate clusters from different categories. The first two lanes are randomly picked PET-3+ and PET-2 clusters. The last three lanes are clusters picked from PET-2, PET-1, and genomic background regions proximal to CpG and promoter with E box. In parentheses are the numbers of validated sites vs. tested sites. (C) Percentages of PET-2 clusters with different binding features positive-validated by ChIP-qPCR.

assay. We verified that 100% (29/29) of PET-3+ and $\approx 47\%$ (9/19) of the PET-2 were real Myc ChIP enrichment events (Fig. 2B). Therefore, we concluded that 593 PET-3+ clusters were genuine Myc binding loci and at least half of the 11,000 PET-2 clusters were resulted from random noise by both statistic analysis and experimental validation.

De Novo Motif Analysis Reveals Enrichment of E Boxes in Myc Binding Regions. With the 593 experimentally determined high quality PET-3+ Myc binding loci, we sought to characterize the properties of Myc–DNA interaction by using the motif discovery algorithm Weeder (20) to detect *de novo* consensus motif for Myc. As expected, 5'-CACGTG-3' was the most prevalent motif found in PET-3+ clusters (Table 1 and supporting information). We also determined the possibility of Myc binding to noncanonical 5'-CACATG-3' E box in the 593 high quality sites and found a statistically significant enrichment of this noncanonical E box ($P = 0$) (Table 1). Furthermore, in 367 (62%) of the 593 binding loci, either one of the E box variants or both were found; consistent with an earlier study that in Myc overexpressed cells, Myc has high affinity to canonical and noncanonical E boxes (12). However, E box-independent mechanisms may also be important factors in dictating Myc binding site selection as there are still $\approx 40\%$ of the binding loci lacking either canonical or noncanonical E box sequence.

Proximity of Myc Binding to Promoters and CpG Islands. Using the 593 highly reliable Myc binding loci, we determined the locations of Myc binding relative to gene coding sequences along the genome. The majority of the Myc binding loci (63%, 372/593) is within a 10-kb range around known gene regions with a strong binding preference toward 5' proximal promoter regions (10 kb upstream of transcription start sites and first introns) (Fig. 3A). Notably, Myc binding in the 5' region is 11-fold higher than its

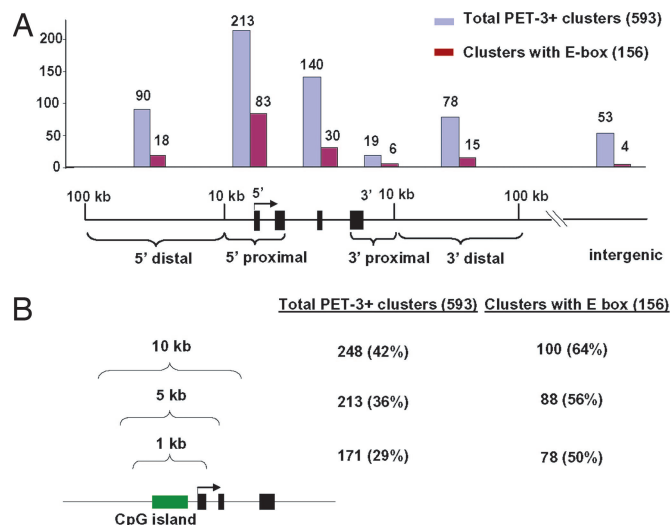


Fig. 3. Locations of Myc binding sites relative to gene structure and CpG island. (A) The locations of 593 reliable Myc binding loci (blue) and 156 E box containing loci (red) were displayed in relation to a gene structure model. The number of binding sites at each particular location is indicated at the top of the bar. (B) Number and percentage of 593 PET-3+ and 156 E box-containing Myc binding loci located within 1, 5, or 10 kb of CpG islands.

binding in 3' region (213/19 for all PET-3+ loci and 83/6 for E-box containing loci). Another predictor of Myc binding is the proximity to CpG islands, which are hypomethylated genomic regions that are frequently associated with transcribed genes (21). We examined the association of the 593 Myc binding loci with CpG regions and found 29% and 36% of total Myc PET-3+ clusters are located within 1 kb and 5 kb of CpG, respectively. When specifically evaluating the subset of 156 PET-3+ clusters containing consensus E boxes, we found more than half of them (88; 56%) were within 5 kb of CpG (Fig. 3B). These results indicate a preference for Myc binding with CpG islands.

Based on the binding characteristics derived from highly reliable PET-3+ clusters; proximity to CpG islands (<5 kb) and proximal promoter regions of known genes (within 10 kb of transcription start site and first intron), and the presence of E-box, are the three most significant features of Myc binding. In total, 326 of the 593 Myc loci determined by PET-3+ clusters exhibit one of these three characteristics (for the list of 593 binding loci and associated genes, see supporting information).

Applying the above criteria to the 11,000 potential Myc binding loci suggested by PET-2 clusters, we found that 263 loci satisfied all three criteria, 1,425 loci have at least two features and 3,703 loci have at least one of the three Myc binding features (1,320 containing E box, 1,854 within 5 kb of CpG islands, and 2,212 in proximal promoter regions) (see supporting information). To estimate the level of true Myc binding to clusters with different features, we randomly picked a minimum of 10 loci from PET-2 clusters with one of any three, two of the three, or all three Myc binding features for ChIP-qPCR validation. As expected, all 14 loci (100%) fulfilling all three criteria are positive (Fig. 2B and supporting information). The percentage of validated clusters from the other six categories varied from 20% (only within 10 kb of promoters) to 58% (with E box and proximity to CpG). Among them, E box containing loci have the highest validation ratio (Fig. 2C and supporting information). These results indicate that much less than 50% of the 11,000 PET-2 sequences are likely to be bona fide Myc binding loci. Next, we sought to estimate how many Myc binding loci were either missed by PET or overlapped with PET singletons. Putative loci that fulfilled all 3 Myc binding features were

identified from PET-singletons (587 loci) or genomic background (3,689 regions) and subjected to ChIP-qPCR testing. Seven of 10 PET-1 sites (70%) and 2 of 11 (18%) genomic background (non-PET) sites were shown to be positive with lower enrichment fold (Fig. 2B), suggesting many potential low affinity binding sites were not captured by PET clusters because of limited sampling. Thus, to ensure the coverage of weaker Myc-DNA interactions, the 3,703 PET-2 loci containing at least one binding features were combined with the 593 PET-3+ derived binding loci to establish a total of 4,296 Myc binding sites identified in this study (see supporting information). Because these 4,296 sites have potentially 50% false positive rate, they were further analyzed by intersecting with gene expression analysis.

Myc-Directed Transcriptome. By using a range of ± 10 kb for proximity, 2,980 genes are associated with the 4,296 Myc binding sites. If the range is extended to ± 100 kb of genes, there are 3,465 genes associated with Myc binding sites. Assuming that the human genome contains 25,000 genes, this would indicate that Myc regulates ≈ 12 –14% of genes directly in B cells. To examine which of these candidate genes are directly responsive to Myc activation, gene expression data from the same cell line (P493) treated with and without tetracycline were obtained by using Affymetrix U133 microarray. Of the 3,465 putative Myc direct target genes, 668 were found differentially modulated in response to Myc by using a cutoff of significance analysis of microarrays $q < 0.05\%$ (22). Of the 668 responsive genes, 406 were up-regulated and 262 were down-regulated by ectopic expression of MYC (see supporting information).

Functional classification of these direct Myc target genes based on Gene Ontology categorization through the PANTHER database (<http://panther.appliedbiosystems.com>) (23) reveals that they are widely distributed among 211 different categories. Many categories from metabolism, cell cycle control, transcription regulation, intracellular signal cascade and biosynthesis are statistically overrepresented ($P < 0.05$) in this 668 genes subset (see supporting information). This distribution is consistent with the view that Myc affects global gene regulatory networks with specific influence on metabolism, cell size increase and cell proliferation.

Myc Target Transcriptional Regulators and Transcriptional Circuitries. Functional classification of 668 direct Myc targets showed nucleic acid metabolism (140 of 668; 21%) as the largest significantly enriched class, of which 49 genes encode transcriptional regulators (see supporting information). Among transcriptional regulator genes directly up-regulated by Myc are *MAX*, *MXI1*, *MXD3*, and *MNT*, which are involved in the Myc/Max/Mad protein network. This observation suggests an additional level of regulation in this protein network, in which Myc and Mad family members heterodimerize with Max exclusively of the other. Other factors function in cell growth control and cell cycle regulation such as *NFκB*, *STAT3*, *ERβ*, *JUN*, *ELK-4*, *CEBP*, and *ETS1* are also found.

Direct miRNA Targets. We found Myc-bound and potentially regulated seven miRNA targets: mir-148, 346, 17, 196, 124, 155, and let-7a (see supporting information). Among them, the mir-17 cluster of miRNAs was shown to be activated in B cell lymphoma and the forced expression of the mir-17 cluster can accelerate Myc-mediated lymphomagenesis in a mouse model (24). The mir-17 cluster was identified as a direct Myc target (25) and two other binding loci associated with mir-148 and let-7a were also validated through ChIP-qPCR (data not shown). Our study suggests that Myc may regulate miRNA targets as means to promote its activity.

Direct Myc Repressed Target Genes. Although many genes have been reported to be repressed by Myc through expression microarray analysis, only a few targets such as *CDKN2B* and *CDKN1A* were previously identified as direct Myc-repressed targets (15, 26, 27). Our global mapping reveals a new group of 262 Myc-down-regulated genes that are bound by Myc. From gene function classification, intracellular signaling cascade, signal transduction and B cell maturation pathways are overrepresented in these direct Myc-repressed genes. Motif search through TRANSFAC revealed that 2 transcription factor binding sites, early B cell factor (EBF) ($P = 4.48E-19$) and ZIN3 ($P = 5.52E-15$), are significantly enriched in the subset of Myc down-regulated genes compared with Myc-induced genes (see supporting information). EBF is a transcription factor required for B cell lymphopoiesis and plays a crucial role in specifying B cell lineage.

Cofactors Collaborating with Myc in Cis-Regulatory Modules. Apart from the obligated binding partner Max, Myc is also known to collaborate with other transcriptional complexes and the transcriptional activation by Myc is modulated through those interactions (28). For example, Myc forms a complex with Miz-1 to repress gene expression (1). In specific instances, Myc may interact with AP-2, C/EBP, HIF-1, Sp1 or Sp3. However, it has not been well delineated what other transcription factors co-regulate their target genes with Myc. We began to decode the cis-regulatory modules responsive to Myc along with other transcription factors, using the highly reliable 593 Myc interaction sites identified by PET-3+ clusters by screening with optimized percentage weight matrix from 1,051 human TF binding sites in TRANSFAC database (version 9.1) (29). Motifs

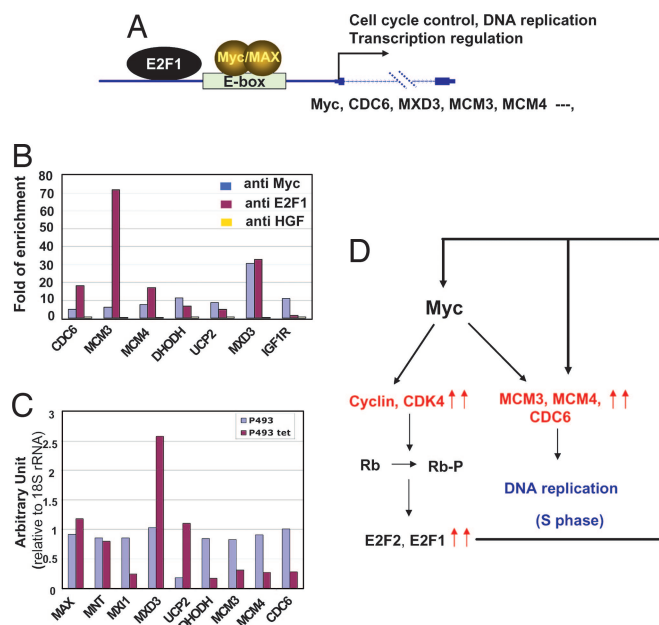


Fig. 4. Myc and E2F1 collaborate to regulate the expression of a number of target genes. (A) Schematic view of Myc-Max and E2F1 binding to the same gene regulatory regions and leading to transcriptional activation of several key cellular processes. (B) Myc ChIP DNA (blue) and E2F1 ChIP DNA (red) from the same P493 cells were subjected to qPCR to validate same binding regions of several target genes. The fold of enrichment is calculated by comparing with negative control HGF ChIP DNA (yellow). (C) Expression of key target genes bound by both Myc and E2F1 were measured by RT-qPCR from P493 cells in the presence (blue) or absence (red) of tetracycline. (D) High Myc expression leads to increased E2F1 activity by up-regulating genes such as cyclins and CDK4. The cooperative binding of Myc and E2F1 followed by transcription activation of key downstream targets leads to the increase of DNA replication and cell cycle progression.

of 20 different TFs were found significantly enriched ($P < 10^{-20}$) from 3- to 24-fold over genomic background (see supporting information). Among these enriched TF motifs, the Myc:Max binding motif was overrepresented with 10-fold enrichment ($P = < 10^{-180}$). Other motifs of known Myc partners such as AP2 and Sp1 were also found. Our analysis of functional categories of genes associated with specific transcription factor consensus sites reveals the possible association of specific gene functions with the binding of Myc and Sp1, AP-2, or MAZ (see supporting information).

The E2F1 motif is specifically enriched 16-fold within Myc binding clusters and 37-fold within the subset of clusters containing E box. When intersected with expression data, 67 of 171 identified loci associated with genes whose expression was modulated in P493 cells (52 were up-regulated and 15 were down-regulated) (see supporting information). Among them, *CDC6*, *DHODH*, *MCM3*, and *MCM4* were confirmed to be bound and induced by both E2F1 and Myc (Fig. 4 B and C). Like Myc, E2F1 also controls cell-cycle progression and DNA replication (30). Thus, deregulation of Myc could potentially lead to uncontrolled cell-cycle progression through a functional link with E2F1 (Fig. 4D).

Discussion

The prevalent role of *MYC* in human tumorigenesis makes the identification of direct Myc target genes critical for our understanding of how Myc contributes to neoplastic transformation. Our global mapping of Myc binding loci by ChIP-PET in a well defined experimental system yields a framework from which functional categories of direct Myc targets appear, cis-regulatory modules emerge, and transcriptional circuitries unfold.

Genomics of Myc Binding Sites. The global Myc binding sites mapped by ChIP-PET identified 4,296 genomic binding loci. The distribution of Myc binding preferentially in gene-rich regions with a bias toward 5' promoters, E box consensus sequence, and CpG islands is consistent with previous more limited studies. The occurrence of Myc consensus E boxes (5'-CACGTG-3') in our refined 4,296 Myc binding loci is $\approx 34\%$, a percentage that is higher than 25% detected from 876 promoters (13) or 756 binding sites from human chromosomes 21 and 22 derived from ChIP-chip data (14). It should be noted that there is no specific enrichment for E box from chromosomes 21 and 22 data because it may contain a significant number of false positives (31). Furthermore, the noncanonical E box is significantly enriched in Myc binding sites and, together with the canonical E box, is present in 60% of Myc binding regions. However, the presence of E box in the binding loci did not correlate with whether or how the associated genes are regulated. In contrast to studies in human cells, whole-genome analysis of *Drosophila* dMyc-responsive genes reveals that $>60\%$ of up-regulated and 90% of down-regulated genes contain a canonical E box within 1,000 bp of the transcription start site (32). Even with the prevalence of E boxes in dMyc target genes, genome-wide mapping of dMyc binding loci reveals significant dMyc occupancy of non-E box binding sites (33). These observations underscore the importance of identifying noncanonical E boxes or non-E boxes to which Myc binds.

The significant overlap of high quality Myc binding loci with CpG islands is consistent with previous studies (12, 15, 34). The fact that Myc preferentially binds to E box motifs within CpG islands suggests that open chromatin is required for Myc to bind its sequence specific target sites or other transcriptional complexes. Myc binding is also shown to be highly correlated with H3K4 methylation and acetylation (31) which further supports an open chromatin conformation is an important for Myc target site selection.

The global mapping of Myc binding loci provides a unique opportunity to discover binding motifs of transcription factors in putative cis regulatory modules. A number of transcription factor binding sites are highly statistically overrepresented in 593 high

confident Myc binding sequences. Interestingly, we found here that sites for E2F1 are overrepresented in Myc direct induced targets, many of which are intimately involved in regulation of DNA replication. Moreover, consistent with a recent study by Cheng *et al.* (35) that ER and Myc interplay in eliciting estrogen response, the estrogen responsive element was also found 7-fold-enriched in our PET-3+ clusters ($P = 5.9E-13$). These new findings, in aggregate, add to the emerging transcriptional circuitry involved in cell replication and growth response.

Myc-Regulated Transcriptional Network. In our analysis, a total of 668 high quality Myc direct responsive genes are associated with expression changes in P493 cells. The functional categories of these 668 genes revealed that a wide variety of different cellular processes like transcriptional regulation, biosynthesis, cell cycle control and signaling transduction are directly regulated by Myc (see supporting information). Substantial numbers of Myc up-regulated genes are involved in pathways to increase activities of protein synthesis and cell metabolism. These include ribosomal proteins, translation factors, RNA polymerase subunits and > 100 genes in TCA cycle, glycolysis, and biosynthesis. These observations support the idea that Myc regulate general protein synthesis machinery and thereby influences cell size control (36, 37).

Most strikingly, transcription factors are among the largest functional category found. In addition to transcriptional circuitries uncovered by binding motifs discovery, direct Myc targets also include members of the Myc/Max/Mad protein-protein interaction network. Although these observations require additional studies, they suggest that in addition to the protein-protein interactions, the Myc/Max/Mad network is also affected by transcriptional regulatory loops.

To date, a small number of genes are shown to be direct Myc repressed targets. The global identification of 262 direct Myc down-regulated targets in our study enables us to dissect the pathways and mechanisms of Myc-dependent gene repression. Among them, 75 genes are in the signal transduction pathways involving MAPK, calcium, Wnt, IGF, TGF- β , phosphatidylinositol, and Jak-STAT signaling pathways. These observations suggest that deregulated MYC could modulate a variety of signal transduction pathways to disseminate oncogenic signals down other paths. Also down-regulated by Myc are the cell adhesion molecules. We also found two transcription factors, EBF and ZIC3, whose motifs are significantly enriched in Myc-repressed genes. EBF is a B cell-specific transcription factor important to cell lineage specification. ZIC3 is a developmental specific zinc finger transcription factor defining early embryo patterning. Down-regulation of EBF and ZIC3 is likely to affect B cell lineage specification; however, the biological significance of these alterations requires further study that is beyond the scope of our current work.

Core Direct Myc Target Genes. One puzzling observation is the limited overlap in Myc binding targets identified here with other published studies. We examined the overlap between our ChIP-PET defined loci and 139 high affinity Myc-bound loci identified by selecting promoter E box within 2 kb flanking the transcription start sites followed by ChIP-qPCR positive in both high and physiological levels of Myc in the same P493 cells (12). Twenty-two sites overlap with our PET-2+ cluster regions and the overlapped sites increased to 47 (34%) when included PET singletons. Likewise, studies in Burkitt's lymphoma cell lines using ChIP-promoter arrays reveal 876 Myc-bound promoters and only 107 of them overlapped with PET-2+ and 386 (44%) overlapped with PET-1+. Moreover, only 132 of 756 (17%) Myc binding sites on chromosome 21 and 22 (30) were overlapped with PET-1+. Systematic validation of 10–15 nonoverlapped binding sites in our system showed 25% to 62% of these sites were indeed missed by our ChIP-PET analysis (see supporting information). Given the apparent low level of ChIP enrichment (14 of 17 positive sites were <5 -fold) and the current

limited PET detection sensitivity, it is not surprising these sites were not found by ChIP-PET. The remaining discrepancy could result from the experimental noise associated with ChIP studies or differences in cell types or specific cell culture conditions. Alternatively, the large differences detected in these studies could be due to the fact that Myc binding are complex and dynamic. Because these studies could be only snapshots of binding events detected by different methods, the real core group of Myc targets would emerge only when we are able to define additional whole-genome Myc direct transcriptomes. Nevertheless, 15 common targets were found that participate transcriptional regulation, DNA replication and cell signaling (see supporting information).

Conclusions and Future Directions. Although novel insights are gleaned from this study and many questions answered, several observations have generated many more questions than those answered. For example, although it is easily conceivable that only less than half of the many Myc-responsive genes are bound by Myc, it remains to be determined why only a small fraction of Myc binding gene loci are associated with gene expression changes. Because it is known that transcriptional factors associate with DNA nonspecifically to scan for specific binding sites, it stands to reason that a snapshot of Myc binding to genomic DNA by ChIP could well capture Myc in the act of promiscuously binding nontarget sites as it contains a nonspecific DNA binding domain (38). This could well account for the large numbers of binding sites for E2F1 and Myc that are well beyond gene regulatory responses (14, 34). Furthermore, because Myc can affect not only transcription at different levels but also the global chromatin structure (39), its binding alone may not be sufficient for gene expression and perhaps its role in these loci is to create poised transcriptional complexes that would be trigger by signal transduction pathways. Finally, because both our ChIP experiments and gene expression analysis were performed on samples taken from a single time point, it is possible that dynamic gene expression changes consequential to Myc binding may have been missed. As such, we envision that the use of customized oligonucleotide tiled arrays covering the high quality ChIP-PET loci

would be a good tool to follow dynamic changes in Myc binding that could then be directly compared with gene expression levels.

Our global mapping of Myc binding loci in a well defined, experimentally tractable system has yielded not only new direct Myc target genes, but it also permits discovery of novel transcriptional regulatory circuitry motifs as well as cis regulatory modules. Future studies emanating from our global mapping of Myc binding sites will provide additional insights into how tumorigenesis is caused by deregulated *MYC*, a prevalent finding in human cancers.

Experimental Procedures

Details. Detailed experimental procedures can be found in the supporting information.

Cell Culture and ChIP. P493-6 cells were cultured, and ChIPs were as previously described (7).

ChIP-PET Experiment. By using the Myc ChIP-enriched DNA fragments, a Myc ChIP-PET library was constructed as previously described (16).

qPCR Assay. qPCR analyses were performed by using ABI PRISM 7900 sequence Detection System and SYBR Green master mix as previously described (17). Two-fold of enrichment was used as cutoff. Primer sequences are in supporting information.

Gene Expression Profiling Analysis Using Microarray. Total RNA isolated from P493 cells in the presence of absence of tetracycline were used to probe Affymetrix U133 Plus 2.0 array following the manufacturer's recommendation as previously described (25).

We thank B. Amati for suggestions and prepublication information. We acknowledge Mr. H. Thoreau and the Genome Technology and Biology Group at the Genome Institute of Singapore for technical support. This work was supported by the Agency for Science, Technology and Research (A*STAR) of Singapore; National Institutes of Health ENCODE Grant 1R01HG003521-01; and National Institutes of Health Grants CA57341 and CA51497. C.V.D. is the Johns Hopkins Family Professor in Oncology Research.

- Adhikary S, Eilers M (2005) *Nat Rev Mol Cell Biol* 6:635–645.
- Grandori C, Cowley SM, James LP, Eisenman RN (2000) *Annu Rev Cell Dev Biol* 16:653–699.
- Blackwood EM, Eisenman RN (1991) *Science* 251:1211–1217.
- Claassen GF, Hann SR (1999) *Oncogene* 18:2925–2933.
- Amati B, Brooks MW, Levy N, Littlewood TD, Evan GI, Land H (1993) *Cell* 72:233–245.
- Cole MD, McMahon SB (1999) *Oncogene* 18:2916–2924.
- Zeller KI, Jegga AG, Aronow BJ, O'Donnell KA, Dang CV (2003) *Genome Biol* 4:R69.
- Watson JD, Oster SK, Shago M, Khosravi F, Penn LZ (2002) *J Biol Chem* 277:36921–36930.
- Remondini D, O'Connell B, Intrator N, Sedivy JM, Neretti N, Castellani GC, Cooper LN (2005) *Proc Natl Acad Sci USA* 102:6902–6906.
- Menssen A, Hermeking H (2002) *Proc Natl Acad Sci USA* 99:6274–6279.
- Solomon MJ, Larsen PL, Varshavsky A (1988) *Cell* 53:937–947.
- Fernandez PC, Frank SR, Wang L, Schroeder M, Liu S, Greene J, Cocito A, Amati B (2003) *Genes Dev* 17:1115–1129.
- Li Z, Van Calcar S, Qu C, Cavenee WK, Zhang MQ, Ren B (2003) *Proc Natl Acad Sci USA* 100:8164–8169.
- Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al. (2004) *Cell* 116:499–509.
- Mao DY, Watson JD, Yan PS, Barsyte-Lovejoy D, Khosravi F, Wong WW, Farnham PJ, Huang TH, Penn LZ (2003) *Curr Biol* 13:882–886.
- Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, et al. (2006) *Cell* 124:207–219.
- Loh YH, Wu Q, Chew JL, Vega VB, Zhang W, Chen X, Bourque G, George J, Leong B, Liu J, et al. (2006) *Nat Genet* 38:431–440.
- Pajic A, Spitkovsky D, Christoph B, Kempkes B, Schuhmacher M, Staegle MS, Briemeier M, Ellwart J, Kohlhuber F, Bornkamm GW, et al. (2000) *Int J Cancer* 87:787–793.
- Kuznetsov VA (2005) in *ISAGE: Current Technologies and Applications*, ed Wang SM (Horizon BioScience, Norwich, U.K.), pp 139–180.
- Pavesi G, Mauri G, Pesole G (2001) *Bioinformatics* 17(Suppl 1):S207–S214.
- Greasley PJ, Bonnard C, Amati B (2000) *Nucleic Acids Res* 28:446–453.
- Tusher VG, Tibshirani R, Chu G (2001) *Proc Natl Acad Sci USA* 98:5116–5121.
- Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremiex O, Campbell MJ, et al. (2005) *Nucleic Acids Res* 33:D284–D288.
- He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM (2005) *Nature* 435:828–833.
- O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT (2005) *Nature* 435:839–843.
- Seoane J, Pouppinot C, Staller P, Schader M, Eilers M, Massague J (2001) *Nat Cell Biol* 3:400–408.
- Seoane J, Le HV, Massague J (2002) *Nature* 419:729–734.
- McMahon SB, Van Buskirk HA, Dugan KA, Copeland TD, Cole MD (1998) *Cell* 94:363–374.
- Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, Meinhardt T, Pruss M, Reuter I, Schacherer F (2000) *Nucleic Acids Res* 28:316–319.
- Sears RC, Nevins JR (2002) *J Biol Chem* 277:11617–11620.
- Guccione E, Martinato F, Finocchiaro G, Luzzi L, Tizzoni L, Dall' Olio V, Zardo G, Nervi C, Bernard L, Amati B (2006) *Nat Cell Biol* 8:764–770.
- Hulf T, Bellosta P, Furrer M, Steiger D, Svensson D, Barbour A, Gallant P (2005) *Mol Cell Biol* 25:3401–3410.
- Orian A, van Steensel B, Delrow J, Bussemaker HJ, Li L, Sawado T, Williams E, Loo LW, Cowley SM, Yost C, et al. (2003) *Genes Dev* 17:1101–1114.
- Bieda M, Xu X, Singer M, Green R, Farnham PJ (2006) *Genome Res* 16:595–605.
- Cheng AS, Jin VX, Fan M, Smith LT, Liyanarachchi S, Yan PS, Leu YW, Chan MW, Plass C, Nephew KP, et al. (2006) *Mol Cell* 21:393–404.
- Johnston LA, Prober DA, Edgar BA, Eisenman RN, Gallant P (1999) *Cell* 98:779–790.
- Grandori C, Gomez-Roman N, Felton-Edkins ZA, Ngouenet C, Galloway DA, Eisenman RN, White RJ (2005) *Nat Cell Biol* 7:311–318.
- Dang CV, van Dam H, Buckmire M, Lee WM (1989) *Mol Cell Biol* 9:2477–2486.
- Knoepfler PS, Zhang XY, Cheng PF, Gafken PR, McMahon SB, Eisenman RN (2006) *EMBO J* 25:2723–2734.